Editorial

Big Data and Parkinson's Disease - from Understanding of Disease to Personalised Medicine

[1]**Yen F. Tai** and [1,2]**Nicola Pavese**

[1]*Division of Brain Sciences, Imperial College London, United Kingdom*
[2]*Institute of Clinical Medicine, Aarhus University, Denmark*

The advancement of information technology in recent years has generated a huge amount of datasets in all areas of scientific research. This has led to new opportunities to harness the 'Big Data' to improve our understanding of diseases - from investigating disease mechanisms to monitoring treatment responses.

The Big Data approach is a data-driven and often hypothesis-free, way of studying a disease. It generally involves multiple research centres pooling together resources and setting up consortia to generate large-scale datasets. Standardised protocols are used to acquire data that traverse molecular, genetic, clinical and imaging domains, hence allowing integration of clinical and biological data for large scale cohort study. Such standardised approach also enables pooled data from different centres to be studied and compared, which greatly increases statistical power. The datasets are usually made accessible to the wider scientific community, further enhancing scientific collaboration and transparency. It also reduces redundant/duplicative efforts and is potentially a more cost effective way to conduct large-scale scientific studies (Editorial, 2014).

Such approach has been particularly productive in the field of genomics. Genome Wide Association Studies (GWAS) involving thousands of patients have successfully identified susceptibility genetic loci for Parkinson's Disease (PD) (Simon-Sanchez *et al*., 2009). In its simplest form, GWAS analysis, as representative of the basic analytical approach to processing Big Data, aims to identify genetic variants such as single-nucleotide polymorphism that distinguish a population with a particular trait or disease from a control population. For diseases with more complex genotypic and phenotypic components such as PD, network analysis has increasingly been used to identify a group or network of interacting genes which may be implicated in disease pathogenesis (Leiserson *et al*., 2013). Such network analysis has also been applied in imaging studies to investigate networks of anatomically dispersed brain regions and their roles in diseases. A number of statistical approaches including linear regression, logistic regression, principle component analysis and latent class analysis have been employed to analyse these large datasets (Wang *et al*., 2014) but it is beyond the scope of this editorial to explain them in detail.

Parkinson's Progression Markers Initiative (PPMI), sponsored by the Michael J Fox Foundation for Parkinson's Research, is a multi-centre collaborative observational study that collects clinical, behavioural, imaging, genetic and biological sampling from cohorts of significant interest - and these include de novo PD patients and participants at high risk of developing PD. It provides a standardised and longitudinal PD database and biorepository which are open to the wider scientific community. It has the stated aim of finding one or more biological markers for the disease as the critical next step towards developing new treatments (PPMI, 2014).

Several high-profile brain imaging collaborative studies including Human Brain Project, Brain Activity Map and BRAIN initiative (Brain Research through Advancing Innovative Neurotechnologies Initiative) have been set up to better understand brain functions, either by creating a large-scale computer simulation of the brain or by establishing a functional connectome of the brain, with the hope that this will eventually lead to a cure for neurodegenerative disorders such as Alzheimer's and Parkinson's disease (Kandel *et al*., 2013).

These large-scale studies can be costly and there are disagreements within the scientific community on how best to run them and whether they will achieve the stated aims (OMECCHBP, 2014). There are also statistical and computing challenges when processing such large volume of data. Many statistical models do not account for possible interdependence of the multiple parameters being sampled and this could lead to reduction of the degree of freedom and violation of some statistical principles. The statistical models themselves may also introduce a degree of bias and false discovery (Wang *et al*., 2014).

Critics have also argued that the Big Data approach can identify correlations between different disease parameters but it does not necessarily establish true causal associations and they struggle to see how this will lead to finding a cure for the diseases in question. Proponents of the Big Data approach counter-argue that such 'signals' are crucial in inspiring more hypothesis-

driven studies to further evaluate the correlations and improve our understanding of diseases (Husain, 2014).

It is important to realise the benefits and limitations of the Big Data approach in studying a disease. It allows a large-scale, unbiased study of data obtained from molecular to individual levels. It does not supplant independent research conducted by individual research groups, which are crucial in generating ideas and hypotheses to complement the Big Data approach. The correlations or 'signals' obtained from the Big Data approach require further targeted studies to elucidate the underlying molecular mechanisms.

Apart from its role in helping us to understand diseases at the population level, the Big Data approach can also be applied, albeit on a smaller scale, on an individual basis to monitor disease fluctuations or treatment responses especially in a disease with marked between- and within-individual variability like PD. As the disease progresses, most PD patients will develop motor complications such as wearing-OFF, 'ON-OFF' fluctuations, dyskinesias and gait freezing. The emergence of these symptoms reflects fluctuations in synaptic dopamine and other neurochemical levels in the brain. PD treatments, especially dopamine replacement therapy, will need to be titrated on an individual basis and, ideally, tailored to specific symptoms at a particular time.

The traditional methods of monitoring fluctuating PD symptoms based on patients' or carers' history, or PD diaries, are laborious and can be misleading as they rely on patients or their carers to accurately identify various 'OFF' or 'ON' symptoms, e.g., tremor versus dyskinesias. Clinic reviews provide only a snapshot of the patients' symptoms and signs in a rather artificial and potentially stressful, setting. There can also be discordance between treatment responses rated by patients and their physicians (Davidson *et al*., 2012).

Mobile or wearable devices worn on patients' limbs or body are being developed to detect 'ON'/'OFF' limb movements, balance deficits and gait disorders in PD patients using specific algorithms. They aim to allow the treating neurologists to monitor an individual's PD symptoms continuously, remotely and objectively in real-life situations (Maetzler *et al*., 2013). Recently, Michael J Fox Foundation announced a collaboration with Intel to use wearable devices to monitor symptoms of PD patients. The devices record more than 300 observations per second from each patient and a Big Data platform has been developed by Intel to analyse the volume of data that will be generated (MJFFPR, 2014). Similarly, Parkinson's UK has also teamed up with Global Kinetics Corporation to provide a wearable device Parkinson's KinetiGraph in a 12-month pilot project (EPDA, 2014).

This type of data with small sample size (individual patient) and high dimensionality (multiple measurements or parameters) is more susceptible to noise accumulation and spurious correlations (Fan *et al*., 2014). One way to get round this problem is by pre-processing the raw data to extract a more manageable secondary dataset of interest (Wang *et al*., 2014). While the statistical methods and algorithms involved might seem complex to most clinicians, the outcome generated is generally user-friendly (e.g., indicating dyskinesias versus bradykinesia) so most clinicians should not require extensive training or IT knowledge to avail themselves of the devices. The more commonly used wearable devices, which are applied on the patients' forearm or wrist, are good at detecting motor fluctuations involving the monitored limb but are less sensitive at detecting gait disturbances or postural instability. To detect these abnormalities, one would often need to apply monitoring devices on patients' trunk or leg and they may be perceived as more intrusive by patients.

Further validation studies are required to verify the accuracy of these devices in detecting or interpreting various clinical parameters before they can be used in routine clinical practice. We also need more evidence to show that such interventions can lead to better patient outcome.

Conclusion

The Big Data approach has the potential to revolutionise medical research by improving standardisation and collaboration across different centres, enhancing statistical power and efficiency of medical studies. It also needs to be complemented by more specific hypotheses-driven studies. On an individual level, such approach can help to better monitor symptoms and lead to personalised treatments for PD patients.

References

Davidson, M.B., D.J. McGhee and C.E. Counsell, 2012. Comparison of patient rated treatment response with measured improvement in Parkinson's disease. J. Neurol. Neurosurg Psychiatry, 83: 1001-1005. DOI: 10.1136/jnnp-2012-302741

Editorial, 2014. Consorting with big science. Nat. Neurosci., 17: 1289-1289. DOI: 10.1038/nn.3830

EPDA, 2014. Parkinson's UK and Global Kinetics Corporation collaborate to provide promising new mhealth technology throughout UK. European Parkinson's Disease Association.

Fan, J., F. Han and H. Liu, 2014. Challenges of big data analysis. Nat. Sci. Rev., 1: 293-314. DOI: 10.1093/nsr/nwt032

Husain, M., 2014. Big data: Could it ever cure Alzheimer's disease? Brain, 137: 2623-2634. DOI: 10.1093/brain/awu245

Kandel, E.R., H. Markram, P.M. Matthews, R. Yuste and C. Koch, 2013. Neuroscience thinks big (and collaboratively). Nat. Rev. Neurosci., 14: 659-64. DOI: 10.1038/nrn3578

Leiserson, M.D., J.V. Eldridge, S. Ramachandran and B.J. Raphael, 2013. Network analysis of GWAS data. Curr. Opin. Genet. Dev., 23: 602-610. DOI: 10.1016/j.gde.2013.09.003

Maetzler, W., J. Domingos, K. Srulijes, J.J. Ferreira and B.R. Bloem, 2013. Quantitative wearable sensors for objective assessment of Parkinson's disease. Mov. Disord., 28: 1628-1637. DOI: 10.1002/mds.25628

MJFFPR, 2014. The Michael J. Fox Foundation and Intel join forces to improve Parkinson's disease monitoring and treatment through advanced technologies. The Michael J. Fox Foundation for Parkinson's Research.

OMECCHBP, 2014. Open Message to the European Commission concerning the Human Brain Project.

PPMI, 2014. Landmark study to find biomarkers. Parkinson's Progression Markers Initiative.

Simon-Sanchez, J., C. Schulte, J.M. Bras, M. Sharma and J.R. Gibbs *et al*., 2009. Genome-wide association study reveals genetic risk underlying Parkinson's disease. Nat. Genet., 41: 1308-1312. DOI: 10.1038/ng.487

Wang, W. and E. Krishnan, 2014. Big data and clinicians: A review on the state of the science. JMIR Med. Inform. DOI: 10.2196/medinform.2913